Semantic Web Overview

"The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

Tim Berners-Lee, James Hendler, and Ora Lassila; Scientific American, 2001

What is Semantic Web?

The first incarnation of the web was primarily for presenting content to the end users. This version of the Web (v1.0) is also known as the "Read" web, where content providers mark up their content using HTML and serve it to the general public for consumption.

The second incarnation of the web, commonly referred to as Web 2.0 is focused on the social interaction. It not only gives content to the public to consume, but also allows the public to participate in sharing their knowledge and experience. This version of the web is also known as the "Read/Write" web.

The third incarnation of the web (Web 3.0) based on Tim Berners-Lee's vision is the Semantic Web. In the Semantic Web, data on the web is made more meaningful through the use of annotations that adhere to some consensual list of well defined terms. These lists can be modeled as dictionary, taxonomy, folksonomy or ontology. These annotations to the published data make the content on the web machine understandable, making it easier to integrate and analyze.

Semantics refers to the meaning behind the terms used to express information or knowledge. By representing that meaning in a standard way, we can come to a shared understanding of very specific concepts or ideas and more easily and explicitly convey that meaning to others.

The "Semantic Web" refers to a set of technologies blessed by the World Wide Web Consortium (W3C) for representing the meaning of information in a standardized way so that others (human and computer) can understand the specific concepts we wish to convey without the inherent ambiguities and confusion of ordinary natural language.

How Can Semantic Web Help NASA Earth Science?

The semantic web empowers the web with access to common sense knowledge that has been previously encoded. A search tool in the semantic web understands the meaning of the entered search terms and provides additional discovery services over and above exact keyword matches. For Earth Science, the semantic web enables automated fusion of data and information in a scientifically accurate way.

Use cases:

- Find data for ozone and ozone-related measurements over the poles
- Find all areas in California in 2009 that had higher than average Sulfur Dioxide emissions

Elements of Semantic Web Technology

- Ontologies provide a shared understanding of concepts
- Ontology Languages (RDFS, OWL, SPARQL, OWL-RL, SKOS) provide for a common framework for representing concepts, relations, and common sense knowledge
- Editors (Protege, SWOOP, CMAP) enable anyone to create and modify ontologies
- Reasoners (Pellet) enable intelligent queries to be posed and solved, based upon the knowledge contained in an ontology
- Storage (Triple stores such as Jena and Sesame) enable large ontologies to be stored and retrieved efficiently
- Integration of components:

Semantic Web layer diagram representing the current layered architecture for which languages, functional capabilities (representation, query, rules) and logic are built upon the Resource Description Framework. At the higher layers are advanced capabilities such as proof and trust. Applications include semantic web services.

State and direction for Semantic Web Technology

Although semantic web technologies can be useful in developing individual applications, their true potential is reached in providing meaningful interoperability among disparate systems on the web. Therefore, a chief goal of the Technology Infusion Working Group is to facilitate the infusion of these technologies into NASA's Earth science data systems. This activity involves assessments and recommendations regarding semantic web technologies in an effort to help Earth science data system developers to incorporate semantic web technology. As more of these systems are placed on a semantic web footing, more of them will be able to communicate with each other (the "network effect").

Today, a diverse set of semantic web technologies is available to the software practitioner. These range from moderately simple languages, such as RDF and OWL, up through advanced reasoners and sophisticated tools for implementing semantic technology. Some of the more basic technologies have reached a state of maturity and stability to allow skilled general software practitioners (without special expertise), to implement systems incorporating significant semantic web technology. Robust commercial and open-source tools are available to support implementation, as well as training and other information resources. Currently, successful implementations usually take place within an application or enterprise, where domain knowledge is easier to assemble and incorporate, though there are notable exceptions, particularly in the area of data search.

On the other hand, some of the more advanced semantic web technologies are still in at a research and development stage, particularly in the area of high-level reasoning, such as geospatial and agent-based reasoning. As such, they also require specialized expertise to implement successfully, and the tools to support these are immature. However, the eventual maturing and adoption of these technologies should eventually enable more sophisticated cross-enterprise applications in Earth science information services, collaboration frameworks and knowledge management.

Perhaps of equal importance is the fact that the field of common ontologies is still nascent in the Earth Sciences domain. Efforts are currently underway to remedy this, such as the SWEET ontology and the ESIP data and services ontology. These common, reusable ontologies are a critical component to bridge the semantic interoperability gaps among disparate systems and users within Earth sciences.

The Semantic Web subgroup has developed a detailed roadmap and gap analysis which outlines which technology and ontology developments in detail are essential for the future development of semantic interoperability in Earth sciences. In addition, the Semantic Web subgroup conducts tutorials and workshops on Semantic Web, with particular applicability to the Earth Sciences. These are designed to help the practitioner in the infusion of semantic web technologies.